By I. A. WARHEIT

A Study of Coordinate Indexing as Applied to U.S. Atomic Energy Commission Reports

Dr. Warheit is head, library department, Argonne National Laboratory, Lemont, Ill.

SHORTLY AFTER it had been proposed that the Uniterm system of Coordinate Indexing be adopted for the research reports of the U. S. Atomic Energy Commission, some of the information officers and librarians concerned with the utilization of AEC research reports undertook a series of independent studies of both a theoretical and practical nature. Upon comparison of the results of these studies the information officers and librarians concerned agreed that a great many problems remained to be resolved, that the system of coordinate indexing requires further development and that in its present state of development, the coordinate indexing system offers no inducement for change, and, indeed, its application to AEC reports would be a retrograde step.

It is the purpose of this paper to summarize the more pertinent of these tests that were applied to the Uniterm System of Coordinate Indexing, omitting, as far as practical, the various theoretical studies.[1] Although the results were negative in the case of the AEC reports, it does not follow that the results will be the same for another body of literature. It is only after a series of tests are applied to various collections[2] that a definite evaluation of Coordinate Indexing as a whole can be made with real assurance. A large body of experimental data is necessary in order to reduce much of the controversy now raging about this novel form of indexing.

One of the many advantages claimed by the advocates of Coordinate Indexing is that the card catalog is radically reduced in size and that there is a similar reduction in the librarian's work load.[3] A comparison therefore was made between an existing AEC reports card catalog for 40,000 documents and its equivalent as a Uniterm catalog.

	Present Catalog	*Uniterm Catalog*
1. Series (Number) Cards	55,000	55,000
2. Author Cards	102,000	102,000
3. Accession Number Cards	0	40,000
4. Subject Cards	180,000	6,000
	337,000	203,000

It should be pointed out that the figure of 6000 Uniterm subject cards is based on claims made by Documentation Inc. as a result of their analysis of the subject headings used by the Technical Information Division (TID) of the Library of Congress

[1] Jack C. Morris, "Evolution or Involution?" *Journal of Cataloging and Classification,* X (July, 1954), 111-18. This was part of the AEC study. *See also* his paper "The Duality Concept in Subject Analysis," *American Documentation,* V (August, 1954), 117-46. *See* especially pp. 138-146. G. E. Randall, "Practicality of Coordinate Indexing," COLLEGE AND RESEARCH LIBRARIES, XV (October, 1954), 417-19.

[2] For another test of Coordinate Indexing *see* C. W. Cleverdon and R. G. Thorne. *A Brief Experiment with the Uniterm System of Co-ordinate Indexing for the Cataloging of Structural Data.* (Royal Aircraft Establishment, Library Memorandum No. 7) January, 1954.

[3] C. D. Gull, "Alphabetic Subject Indexes and Coordinate Indexes: an Experimental Comparison," COLLEGE AND RESEARCH LIBRARIES, XIV (July, 1953), 276-81; also as chapter 6, p. 56 of Mortimer Taube & Associates, *Studies in Coordinate Indexing.* (Washington: Documentation, Inc., 1953).

and of the Document Service Center (DSC) in Dayton.[4] A careful count of the AEC subject heading list showed that it would take almost 8500 Uniterms to cover all the subjects adequately.[5] But in order to reduce the areas of controversy, Documentation Inc. figures are used wherever possible. Also, in this instance, the difference of a few thousand cards will not materially affect the results.

These 6000 Uniterm cards would carry 342,000 postings. This is based on 8.55 Uniterms per title. Although Gull claims "that the average of 6.88 unit terms per report is the optimum for Coordinate Indexing of . . . TID reports," the figure of 8.55 Uniterms per title is based on a tabulation of 311 ASTIA cards which carry Uniterm tracings. Actually the number of postings would be higher in an AEC catalog. Just as there are about twice as many subject tracings on an AEC catalog card as compared to an ASTIA card, the coordinate indexing of one sample of 200 AEC reports required 11.41 Uniterms per title. Thus one should expect some 456,400 postings, but again the lower Documentation Inc. figures will be used.

On the basis of past experience and a series of test postings, it was determined that, day-in, day-out, the average clerk could sort, mark and file at the rate of 60 cards an hour and hand post at a rate of 40 accession numbers an hour. The posting rate may seem rather low, but it must be remembered that Uniterm tracings cannot be presorted but must be posted in accession order.[6] Therefore the whole index must be worked through for each title separately. If more than one posting clerk is necessary then a certain type of presorting is possible. The entries could be grouped by their last digit and then posted. This would still mean working through the alphabet for each entry but would permit more than one person to do the work at the same time. So, even allowing for the development of mechanical devices to speed the pulling of cards and the posting of numbers, it is hardly possible, over a long period of time, to average much more than 40 postings per hour. On this basis, (the AEC catalog would require 5616 hours of work, but the Uniterm catalog would require 3383 hours for filing and marking plus 8550 hours of posting, for a grand total of 11,900 hours. Reducing these to more understandable figures, we find that, for every 100 titles received,

the regular AEC catalog requires:

843 cards	=	14.05 hours
0 postings	=	0 hours
		14.05 hours

the Uniterm catalog requires:

508 cards	=	8.45 hours
855 postings	=	21.37 hours
		29.82 hours

Thus, converting AEC libraries to Coordinate Indexing would mean at least a doubling in catalog maintenance costs. Just to break even it would be necessary to pull, post, and refile 855 items in 5 hours and 36 minutes, or over 150 items an hour! It should also be pointed out that it would be very difficult for more than one person at a time to work at a Uniterm catalog. Since

[4] Richard B. Thomas and C. D. Gull. *The Choice of Uniterms for a Coordinate Index of Scientific and Technical Reports,* Technical Report no. 6 . . . for the Armed Services Technical Information Agency, March, 1953; also as chapter 5, p. 47, Taube, *op. cit.* Combining the TID and DSC subject heading lists, Documentation Inc. converted them into a list of 6582 Uniterms.

[5] The AEC subject heading list has some 10,000 subject headings and a little over 1500 *see* references. The 8500 Uniterms would include 1117 names of organic compounds and 1406 proper geographic names used for geology reports. The total actual reduction in subject headings if Uniterms were used would be in the order of 40%. The figure will vary from 36% to 53% depending on the inclusion or exclusion of the geology and organic compounds headings.

[6] Using high speed Burroughs or National Cash Register machine, the average posting clerk in a bank can post from 400 to 600 accounts each day. This includes the necessary preparation work such as alphabetization of checks and deposit slips, but it does not include verification of items.

posting would be a full-time job, it would be necessary either to have two Uniterm subject catalogs, each of which could be posted on alternate days with one thus always available for the public, or to have the posting clerks work nights. On the other hand the fact that the Coordinate Index requires fewer cards, does mean a slight reduction in printing costs.

More important than cost, however, is the efficiency of the system. Can Coordinate Indexing give better retrieval than the regular catalog? To test this, three librarians at an AEC laboratory Uniterm indexed some 500 of their reports in accordance with the instructions in the *Installation Manual for the Uniterm System of Coordinate Indexing* prepared by Documentation Inc. and published by the Document Service Center of the Armed Services Technical Information Agency, October, 1953. The report numbers, used in lieu of accession numbers in order to simplify the task, were then posted to the appropriate Uniterm cards. The posting was done by hand by two people, one pulling and refiling the cards, and the other posting. It soon turned out that the two posting clerks were unable to keep up with the three indexers. Therefore, while some 500 reports were actually indexed, only 200 were ever posted. As a result, the test was confined to the 200 titles which could be posted.

When the sample catalog was completed, other librarians and "scientist users" experimented with it. All were sufficiently well acquainted with the subject matter of the reports included in the sample to select reference questions capable of answer by the material indexed. Almost as soon as the tests began, four things became apparent. First, there were an inordinately large number of false drops. Second, many items could not be retrieved. Third, the absence of any descriptive information or abstract made many of the searchers feel they were hunting blind and brought immediate protests. Fourth, certain key Uniterms were required in a large percentage of the searches and since these cards usually had the most entries their use was often the slowest. As a result, usually only one or, at the most, two searches could be conducted simultaneously at the index. The fact that the absence of one Uniterm card from the file nullified much of the usefulness of the entire index also made the prospect of losing a Uniterm card frightening.

A careful analysis of the two hundred test reports was then undertaken to see why the difficulties were experienced and if they could be corrected. Twenty-three reports or 11.5% of the total were adequately indexed. All of these were concerned with concrete things: design, fabrication, testing, etc. None dealt with theoretical subjects. Thirty-nine reports or 19.5% of the total could be salvaged if the reports were divided up into separately numbered sections so as to avoid the many false drops. These 39 reports had to be separated into 334 sections. Thus the 200 reports under study actually represented 495 units of work, a doubling in the effective size of the report collection to be indexed.

Thirty-five reports (17.5% of the 200) were made retrievable by exercising much tighter control of the Uniterms by such means as elimination of synonyms, the addition of cross references, definitions, qualifying phrases and that whole apparatus of subject heading control which Coordinate Indexing claims to eliminate. Difficulty was experienced with 54 reports (27%) because the coordination of multiple related Uniterms pertinent to the subject matter of a document could not be varied. That is to say, if four Uniterms in a certain combination were necessary for the retrieval of a report, variations of these Uniterms, that is using two or three of the terms or changing their order gave meanings which bore

no relationship to the material sought. In other words, the mere accretion of terms did not always make an idea more specific but rather changed meanings. Yet it was necessary to leave the terms "free" since *some* of the word combinations did fit the subject matter. This problem tended to occur with highly technical subjects and could be solved only by providing elaborate definitions and explanations limiting the meanings of the Uniterms especially when used in certain combinations. These warnings and guides were far more complicated than the usual explanatory statements found in conventional subject heading lists for the same subject matter.

Almost all the reports showed instances of miscoordination of adjectives, sub-headings and compound terms. For example, a particular chemical report, involving eleven elements, required the use of three adjectival Uniterms, two sub-heading-type terms, and one other term, all of which could apply to all the elements. Each of these actually applied to only one or two of the eleven elements, but could be coordinated with all. Yet it was not practical to subdivide the report into sections to prevent the many false drops. Thirty-eight reports (19%) had difficulties because adjective Uniterms could be misapplied. Sub-headings, especially common ones like "design," "properties," and "calculations," caused trouble with 79 reports (39.5%). Compound subject headings split into Uniterms led to false drops in 144 reports. (72%).

This whole question about the false drops, or as it is sometimes referred to, the "noise" in this information scheme has been the subject of some of the most heated debates about Uniterm indexing. The proponents of Coordinate Indexing admit that there are false drops but that these occur infrequently and are so widely scattered and statistically insignificant that they can be neglected. On the other hand, some theoretical studies of the permutations and combinations possible with a group of Uniterms have shown that extremely large numbers of false drops could be expected. It has been amusing to note that the same mathematics used by Documentation Inc. to illustrate the difficulties catalogers experience in selecting the proper permutations of multiple-term descriptions in an alphabetic index, is also used in the theoretical studies to demonstrate the number of false drops to be expected with Coordinate Indexing.[7]

It is obvious, of course, that the number of false drops which may be experienced will vary greatly from sample to sample. For example, reports whose subject matter is widely scattered will produce very few false drops, whereas reports which are concentrated in extremely narrow subject fields will produce many false drops. Similarly, simple short documents whose information is confined to two or three concepts which can be described in a few words will produce very little "noise" in an information system, but long treatises covering many very technical topics, which require many Uniterms to describe them adequately, will produce a deafening roar. It is obvious then that each collection of information will cause its own characteristic number of false drops if Uniterm indexing is used. Therefore, at one of the AEC laboratories a group of frequently used subjects was analyzed to see how many false drops Coordinate Indexing would produce.

Since uranium corrodes very readily, the usual practice is to can it or clad it with other metals. There is, as a result, an

[7] Mortimer Taube, C. D. Gull and Irma S. Wachtel. *Unit Terms in Coordinate Indexing,* Technical Report no. 3 . . . for the Armed Services Technical Information Agency, November, 1952; also as chapter 4, p. 37, Taube, *op. cit.*

As an unpublished appendix to his "Evolution or Involution?", Mr. Morris analyzed an artificial group of 36 reports all of which could be covered by 5 Uniterms. Setting up all the possible intelligible combinations he could with these 5 Uniterms, Mr. Morris was able to "regenerate" 92 valid references (39% of the total) 81 false drops (34% of the total) 27 "confused" concepts (11%) and 39 "far-fetched" concepts (16%).

extensive literature in the AEC libraries on the corrosion of uranium, the corrosion of cladding metals, the compatibility of uranium with these metals, and so on. A report on the corrosion resistance of a cladding metal, therefore, often contains additional information on various properties of uranium, but no information on the corrosion of the uranium. At the time of the Coordinate index test there were 175 valid references in the catalog to the corrosion of uranium. However, a count of the corrosion references to a series of metals produced the following number of false leads to uranium corrosion:

	Number of cards in catalog	*No. of Uniterm false leads*
ANTIBODIES—		
Radiation Effects	11	6
ENZYMES—		
Radiation Effects	51	73
METABOLISM—		
Radiation Effects	29	62
NUCLEIC ACIDS—		
Radiation Effects	10	26
	101	167

ALUMINUM—CORROSION (300) cards produced 102 false leads
ZIRCONIUM—CORROSION (200) cards produced 44 false leads
STAINLESS STEEL—CORROSION (250) cards produced 70 false leads

Total false leads to URANIUM—CORROSION = 216

Actual number of entries for URANIUM—CORROSION = 175

In other words, three subject areas with a total of some 750 references produced 216 false leads to URANIUM-CORROSION. And this did not exhaust the possibilities, for the corrosion of all the other metals: copper, tin, zinc, etc. would also have produced their quota of false drops. Thus, in this instance, the false drops would exceed the valid references.

On the other hand, a check of a subheading applied to alpha particles produced only a little over 2% of false leads to gamma radiation. Conversely, seven reports on the alpha emission of americium 241 produced three false leads to tritium, three to neptunium and one each to isotopes of cesium, uranium, cobalt, fluorine, niobium, argon and to deuterium.

Tests were also made on a series of biological subject headings using a reverse approach; namely how many false leads would sample headings produce using the same sub-heading? These false drops would, of course, be scattered throughout the catalog and they would be harmful only if there was an accumulation at any one point. Nevertheless, they give some idea of the ratio of false drops to valid leads one might expect.

Even allowing for duplications, the number of false leads thus produced would be intolerable.

It is true that the subjects chosen made for a very severe test of Uniterms. They were not average samples. However, they were not chosen because they would produce the largest number of false drops. They were chosen because they are vital to the work of the AEC laboratories and it is in these subject areas, and others like them, that the libraries must produce the necessary information. If Coordinate Indexing cannot help here, then it has little to offer the AEC libraries.

Could these false drops be eliminated by various devices? In part, yes. They can be mitigated somewhat by segmenting certain reports and numbering the sections separately. In one test group, almost 20% of the reports thus could be salvaged. This, however, materially increased the cataloging, filing and posting work loads. Another solution was to use polyterms. This means

essentially abandoning Uniterms and using full-scale subject headings with none of the advantages of the card catalog. Again this would have to be applied to those very subject areas which are used most frequently and are of the greatest importance, leaving Uniterms for the less used materials which can be controlled adequately by the simplest cataloging. In spite of all efforts to modify the indexing in order to prevent false drops, 45 reports in a sample of 200 still produced false drops.[8]

In addition to false drops, several of the tests were troubled with the problem of "lost" information. In other words, coordination of Uniterms failed to produce information which was known to exist. In one test 16.5% of the reports proved irretrievable.[9] There seemed to be three reasons for this: 1. All the essential aspects of the report were not covered by the title, section headings, and abstract. Where the indexing was, in accordance with the Uniterm installation manual, confined to the title and abstract, in over 16% of the cases the indexers had to go back and carefully read the report and re-index it. 2. Occasionally the various levels used different terms instead of coordinated Uniterms. This would happen with reports that covered individual components of a larger apparatus. Since the name of the larger apparatus did not occur in the title or abstract, a Uniterm for it was not provided. Again, this was solved by applying standard cataloging practices and supplying Uniterms even though they did not appear prominently in the report.[10] 3. The third reason was a psychological one. People are accustomed to certain patterns for subject headings and they do not think to look under such strange terms as HIGH, HALF, ULTRA, 90, B, RUBBING, VERY, H, 1952 and so on. There is also a fourth reason: some people do not formulate their questions very precisely, but close in on their desired references by trying out various subject headings. This form of browsing and relying on suggestions is practical in an ordinary catalog where titles and abstracts are available, but cannot be done in a Uniterm catalog.

Direction of action between coordinated Uniterms also caused difficulties. In one test sample, 12% of the reports indexed demonstrated this difficulty. These were primarily chemical reports involving possible reversible chemical reactions and physics reports including particle reactions. The solution for this was to use polyterms or "bound" terms, which is just another way of saying ordinary subject headings.

In addition, some minor problems which could lead to difficulties showed up in the tests. The visual coordination of long columns of numbers was fatiguing and caused many errors. Some means of machine coordination is necessary for extensive searches. More mechanization must also be developed to speed up posting. Unlike filing mistakes, posting errors were extremely difficult to detect and caused trouble.[11] Users objected to the lack of selectivity. A person wanting a general paper on an entire field is also burdened with references to all the specific papers on particular aspects of the field. The only way to cull this mass is to take the extra step of checking the accession cards.

---

[8] In the British tests, Royal Aircraft Establishment, *op. cit.*, there was, in general, one false drop for each relevant item found.

[9] The British test, Royal Aircraft Establishment, *op. cit.*, reported 15% lost items. However, using their U.D.C. catalog, they lost 50% of the searched items!

[10] As examples of this, taken from ASTIA cards, one must use both Uniterms FILAMENTS and CATHODES to find all the filament references; one must search CRYSTALS, CRYSTALLOGRAPHY, QUARTZ and many other headings to find the necessary crystal references. Of course such difficulties can be overcome by providing cross-references, but it does illustrate the high degree of scattering that the Uniterms induce.

[11] It may be necessary to adopt some verification system as used in bank posting or in punch card work. This would involve either checking a carbon tape or duplicate posting. In either case it would mean a doubling of posting time.

On the positive side of the ledger, Coordinate Indexing did save one-third of the catalog space. This would mean the replacement of a 96 drawer 3 x 5 card file by nine or so 5 x 8 drawers. The actual indexing went very fast, provided the indexer did not go beyond the title, section headings and abstract. This, however, often led to inadequate indexing. In actual practice, more time would have to be spent by the indexer reading the report. Still, it is believed the indexing would be faster than routine cataloging.

It could not be determined if Coordinate Indexing could help retrieve more information by its more flexible specificity and through its increased number of access points. This was due to the fact that the librarians were working with fixed samples where the known, against which the unknown was being measured, was derived from the standard catalog.

Because of the very negative results of these preliminary experiments, the AEC librarians decided to suspend any further work on Uniterm Indexing and to await the results of field trials now being made at the College of Aeronautics, Cranfield, England, and with ASTIA cards in this country.

A Reply to Dr. Warheit

There are a great many points in Dr. Warheit's paper to which we might take exception, but there are only four specific matters which urgently require clarification.

1. Although the initial statement of Dr. Warheit's position refers to a "series of studies" of both a theoretical and practical nature and although Dr. Warheit indicates that his paper is based to a large extent on actual tests rather than on theoretical considerations, most of the figures in the paper are based on a single test in which only 200 items were indexed and posted. There is presented at the end of these remarks the tabulations made from that test, and it can be noted how many of Dr. Warheit's figures are taken from this tabulation.

2. The original Los Alamos test concluded that Uniterm indexing was inadequate indexing leading to misconception, partial retrieval, false drops, etc. This conclusion is reiterated by Dr. Warheit. In describing the conditions of the Los Alamos test, those who conducted the test stated:

> In this initial study, Taube's rules were followed as closely as possible:
>
> > Uniterms were assigned from the titles, abstracts, contents lists, and occasionally paragraph headings. The reports were not read.
> >
> > .....
> >
> > The problems of synonyms and homonyms were considered of little importance and were generally disregarded.

No references are given for these amazing statements.

No such rules or anything like them can be found in anything we have written, nor have we been able to trace the source of this misconception. In Dr. Warheit's paper there is a reference to the *Installation Manual* (no page citation) as the source of the instruction to index reports without going beyond the title and abstract. We have gone over the *Installation Manual,* word by word, trying to pin down the source. Although there is an explanation of the method of setting up a Uniterm Index based on standard cards supplied by other indexing organizations (the title and abstract on a card can be used if they give more information than the subject-headings), it is difficult to believe that Dr. Warheit and the Los Alamos people could have confused the indexing of a document with the conversion of a completed index to Uniterm form. Yet, this is the most charitable explanation we can devise for an error which is as gross as it is gratuitous. It is certainly not necessary to conduct tests to arrive at the conclusion that sloppy and partial indexing makes bad indexes.

3. No valid conclusion about posting can be drawn from handling 200 reports which required the creation of 910 Uniterm cards. As the vocabulary stabilizes (and neither Dr. Warheit nor the Los Alamos testers deny that this will happen) the rate of posting even under the worst conditions will increase by

several hundred per cent. There really was no need for Dr. Warheit to base his conclusions on such a small sample. The people in charge of information at AEC headquarters in Washington have inspected an installation in which close to 50,000 items have been posted. They know that cards can be presorted and that postings can be accumulated on punched cards. Even with a simple manual system, different posters can post different parts of the alphabet; and a well tabbed manual system has enabled us to average between two and three postings per minute, i.e., 120 to 180 per hour as contrasted with Dr. Warheit's figure of 40.

4. It is a recognized limitation of the Uniterm System or any machine system of coordinate indexing that it does not readily permit browsing in a subject arrangement of titles and abstracts. Dr. Warheit's remarks on this point may serve to bring it to the attention of those who missed it in our report on *The Evaluation of Information Systems*. But here again it seems that his comments cannot be based on any actual tests because AEC security regulations do not permit readers to browse through the catalog. The contribution which the Uniterm System makes to improved security would counsel its adoption by the AEC above all other agencies, even if it were only barely as effective as the present system. Hence, I feel that it is unfortunate that the AEC librarians see nothing good in the system. Actually, the only "controversy now raging about this novel form of indexing" and up to now the only completely unfavorable published articles on the Uniterm System are products of AEC librarians.—*Mortimer Taube, Documentation, Inc.*

Tabulation of an Evaluation of Coordinate Indexing for 200 LA Reports

	Reports	% of total Reports*	Uniterms Involved	% of total Uniterms*
Adequately indexed	23	11.5	—	—
Required modification:	177	88.5	—	—
Need for report division	39	19.5	—	—
Need for better Uniterm selection	35	17.5	84	3.68
Problem of scientific terminology	54	27.0	108	4.74
Need for polyterms	—	—	1344	58.82
Adjectival	38	19.0	164	7.17
Sub-heading-type	79	39.5	184	8.05
Complex headings	144	72.0	996	43.60
Problem of direction	24	12.0	44	1.93
Adequately index after modification	99	49.5	—	—
Still presenting problems of specificity after modification	45	22.5	—	—
Irretrievable after modification	33	16.5	—	—

* Total number of reports indexed = 200
Total number of uniterms used = 910
Total number of times uniterms were used = 2283

Selected Reference Books of 1954-55

(Continued from page 277)

the 1940 edition (*Guide* V356) of this critical and selective bibliography to include material—books and articles—written since 1940 concerning the humanities and the social sciences in Japan. The same classified arrangement has been used, but an entirely new section on World War II and the Occupation, 1941-52, has been added. Full bibliographical details are given for each item, and brief descriptive notes for some. There is an index to titles and personal names. Full titles of abbreviated periodicals, are listed under "Periodicals" in Chapter III.—M.C.

Cordier, Henri. Bibliotheca sinica

———. *Author Index to the Bibliotheca sinica of Henri Cordier.* (2d. edition, 4 vols. Paris, 1904-08. Supplement, 1 vol., Paris, 1924). Compiled, issued and distributed by the East Asiatic Library, Columbia University Libraries. N.Y., 1953. 84p. $3.25.

For Cordier see *Guide* V191. This work provides a much needed index to this valuable bibliography.